Tracking Multiple Objects in 3D

João P. Barreto
Inst. Systems and Robotics
Dep. of Electrical Eng.
Coimbra 3030
PORTUGAL

Paulo Peixoto
Inst. Systems and Robotics
Dep. of Electrical Eng.
Coimbra 3030
PORTUGAL

Jorge Batista
Inst. Systems and Robotics
Dep. of Electrical Eng.
Coimbra 3030
PORTUGAL

Helder Araujo
Inst. Systems and Robotics
Dep. of Electrical Eng.
Coimbra 3030
PORTUGAL

Abstract

In this paper a system for tracking multiple targets in 3D is described. The system is made up of two pan-and-tilt units that are attached to the extremities of a rotating arm. This configuration has several advantages and can deal with several specific instances of tracking more than one target. A control strategy that guarantees equal target disparities in both images whenever targets are seen by both cameras is presented. This has advantages for segmentation and trajectory reconstruction. Target images are simultaneously visible in the cameras, enabling the recovery of the targets 3D trajectory. It is also shown that mutual occlusion occurs in a well-defined configuration and can therefore be dealt with.

1 Introduction

Surveillance applications are extremely varied and many of them have been subject of a significant research effort [1, 2, 3, 4, 5]. Besides the applications themselves, the techniques employed include a wide range of computer vision domains. For example high-level modeling and reasoning [3, 4] is extremely important for the development of the applications since interpretation of the events is crucial for the usefulness of automated surveillance. Motion segmentation is another issue with high relevance for surveillance applications. There is a wide range of approaches for this problem using different techniques and assumptions.

In many applications visual attention must be focused in a certain moving target. Tracking behaviors can be implemented to achieve this goal. The idea is to use a rotational (pan and tilt) camera to keep the target projection in the same image position (usually its center). If the same moving target is simultaneously tracked by two or more cameras with different positions, it is possible to recover the 3D target trajectory. The recovery of the 3D target trajectory can be done by using only image data or by combining image information with the changes in the system geometry that occur as a result of the tracking process. These changes in geometry are typically measured using motor encoders.

In previous work two cameras were used to track and recover the 3D trajectory of a single moving human target. By combining this system with a third fixed wide-angle camera, it was possible to select (based on their relative positions) one among several non-rigid targets, to be tracked binocularly (and therefore in 3D) [6].

In security applications in environments with multiple moving targets, attention must be focused in two or more targets.. Two independent rotational cameras can be used to track two different targets. By combining a third fixed wide-angle camera, it is still possible to select (based on high-level reasoning) two among several moving targets. However the implementation of the tracking behavior presents additional dificulties, in particular when both targets are simultaneously visible by the moving cameras. In this case motion segmentation in image is needed and mutual occlusion may occur. In this paper we discuss strategies to deal with these specific instances of tracking of multiple targets. A system to efficiently track two independently moving targets using two cameras is pre-

sented. The recovery of 3D target trajectories is possible whenever both targets are visible by both cameras. Our configuration guarantees that if both targets are visible by one camera then they are also visible by the other.

2 The vision system

Let us assume a target as an independent source of motion. It can be an isolated object or a set of objects with motion restrictions between them. Thus, one pan-tilt camera can track at most one independent target. Our goal is to track two independent targets using no more than two cameras, and recover the 3D motion whenever the targets are visible from both cameras.

One obvious solution is to use two pan-tilt units placed on different locations. One of the units, for example unit1, will be tracking one of the targets, target1, and the other unit (unit2) tracks the other target (target2). Target1 3D trajectory can be recovered whenever the target is in the field of view of unit2. The same happens for target2 and unit1. One problem in this configuration is mutual target occlusion. Occlusion is a serious problem from the point of view of both tracking and 3D trajectory recovery (since the 3D trajectory recovery requires that both targets are simultaneously visible in both cameras). One solution might be to use several pan-tilt units so that all points of the area to be surveyed are always visible in at least two cameras. Increasing the number of cameras beyond the strict minimum required adds complexity to the system and, for a limited number of targets, it is not an economical solution.

Using two fixed pan-tilt units is a solution with several limitations. With such a configuration target mutual occlusion can not be avoided. In several circumstances the system does not work properly. In this paper we discuss the use of an additional degree of freedom to change the position of the two pan and tilt units during operation. The system mobility is increased and the interaction ability with the scene is extended. These new features are used to control the position in each of the images of the target that is not being centrally tracked by the corresponding camera. Thus, occlusion can be controlled, and tracking and 3D trajectory recovery become more robust.

2.1 Two targets in the field of view of a rotary camera

$$x_{img}^{L2}(\alpha_p) = f.\frac{x_2z_1 - x_1z_2 + B\cos(\alpha_p)(z_1 - z_2) + B\sin(\alpha_p)(x_1 - x_2)}{z_1z_2 + x_1x_2 + B^2 + B\cos(\alpha_p)(x_1 + x_2) - B\sin(\alpha_p)(z_1 + z_2)} \quad (1)$$

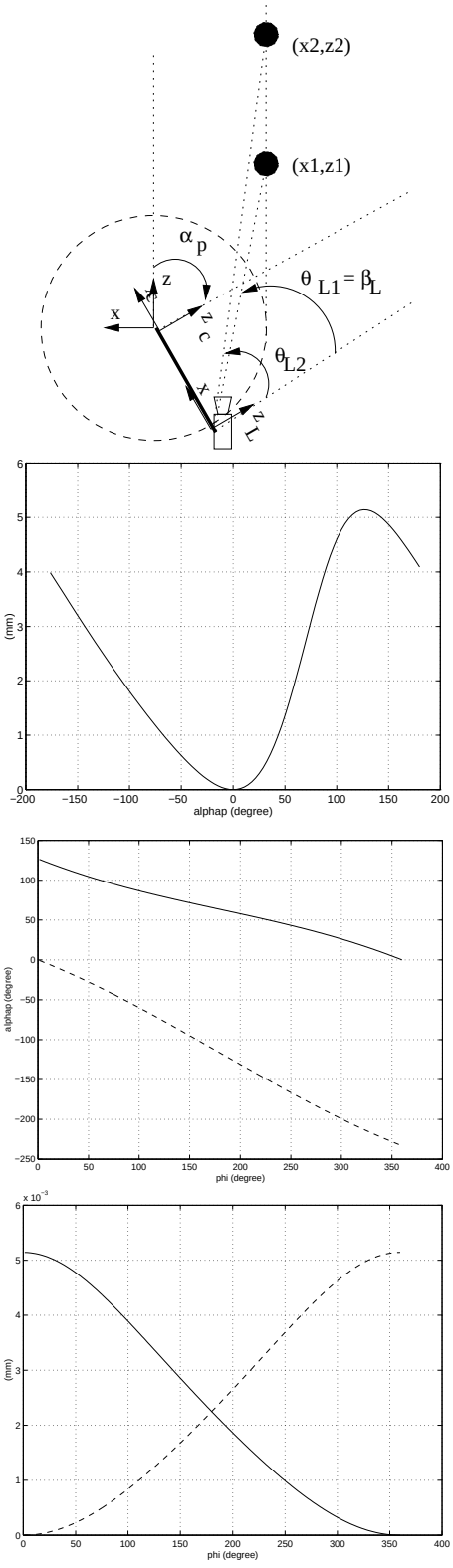

Figure 1: 1:One camera on a circular path. 2: $Target_2$ position in the image (one camera situation). 3: Solutions for α_p that guarantee equal disparity (as a function of ϕ). 4: Disparity in the image (as a function of ϕ

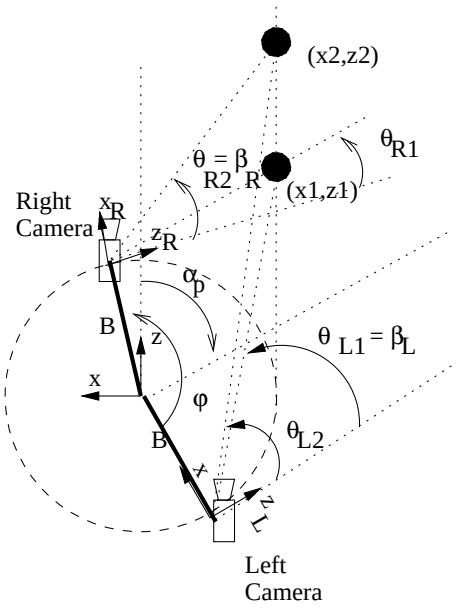

Figure 2: The vision system with the two pan-tilt cameras placed in symmetric positions on a circular path

Assume a camera with pan-tilt degrees of freedom that can be placed along a circular path. Fig. 1(1) depicts a rotary arm of length B with the camera at its tip and two non-moving targets outside the area circumscribed by the arm. Assume the camera verged in $target_1$ ($\beta = \theta_{L1}$). The $target_2$ position in the image is given by equation 1 (where f is the camera focal length,(x_1, z_1) and (x_2, z_2) the targets coordinates, and α_p the arm rotation angle).

Whenever the target projections in the image are too close the performance of segmentation algorithms decreases as well as the accuracy of velocity and position estimation. Therefore, maximization of the distance between the target projections increases the robustness of visual processing. This goal can be achieved by controlling the arm rotation angle α_p. Fig. 1(2) shows the $target_2$ position in the image as a function of α_p for the situation depicted in Fig. 1(1). $Target_1$ is always in the center of the image because the camera is verging on it. This function is periodic (with period 2π) and has one maximum and one minimum in each period. Notice that for $\alpha_p = 0$ occlusion occurs. The optimal rotation angle is the α_p value that maximizes the disparity between the two targets in the same image (computed by solving $\frac{dx_{img}^{L2}}{dt} = 0$).

2.2 Two targets in the field of view of two rotary cameras

$$-x_{img}^{R1}(\alpha_p, \phi) = f.\frac{x_2 z_1 - x_1 z_2 + B\cos(\alpha_p+\phi)(z_1-z_2) + B\sin(\alpha_p+\phi)(x_1-x_2)}{z_1 z_2 + x_1 x_2 + B^2 + B\cos(\alpha_p+\phi)(x_1+x_2) - B\sin(\alpha_p+\phi)(z_1+z_2)} \quad (2)$$

The 3D coordinates of both targets are needed to control the arm rotation. The targets positions can only be estimated by using a second camera. Our goal is to simultaneously track the two targets. Therefore this second camera must also have pan and tilt degrees of freedom. Consider the camera placed at the extremity of an arm with length B. The arm rotation angle is $\alpha_p + \phi$ where ϕ is the angle between the two arms (see Fig. 2). This second camera is verged on $target_2$. The $target_1$ position in the image is given by equation 2. The arm angular position that maximizes the disparity between the two targets images is the same for both cameras. Therefore this criterion leads to a situation where the two cameras are placed in the same position on the circular path, one fixating $target_1$ and the other fixating $target_2$. However the two optical centers can not be simultaneously at the same point. In other words, ϕ must always be different from zero.

$$K_w * \sin(\phi) = (K_x(1-\cos(\phi)) + K_z\sin(\phi))sin(\alpha_p) + (K_z(1-\cos(\phi)) - K_x\sin(\phi))\cos(\alpha_p) \quad (3)$$

$$\begin{aligned} K_x &= (\rho_1^2 - B^2)x_2 - (\rho_2^2 - B^2)x_1 \\ K_z &= (\rho_1^2 - B^2)z_2 - (\rho_2^2 - B^2)z_1 \\ K_w &= B(\rho_1^2 - \rho_2^2) \\ \rho_1 &= \sqrt{x_1^2 + z_1^2} \\ \rho_2 &= \sqrt{x_2^2 + z_2^2} \end{aligned}$$

An alternative criterion to control α_p and ϕ is to keep equal disparities between the target projections in both cameras. The symmetry is useful to increase the robustness of the visual processing and guarantees that whenever the two targets are seen by one of the cameras the same happens for the other (this is important for 3D motion/trajectory recovery). Equation 3 gives the relationship between α_p and ϕ that has to be satisfied to achieve equal disparities in both cameras. This expression is derived by using $x_{img}^{L2} = -x_{img}^{R1}$ (see equations 1 and 2).

Given both target coordinates, equation 3 can be used to compute the left arm angular position α_p that guarantees equal disparities in both cameras as a function of the angular difference ϕ between the two arms. As can be observed in Fig. 1(3) for each ϕ there are

two solutions for α_p that correspond to different disparities in the image. Notice that the maximum and minimum values of the achievable disparity range are obtained for $\phi = 0$ and there are two solutions (ϕ, α_p) that give the same disparities.

Thus we can guarantee equal disparity in both cameras by relating α_p with ϕ according to equation 3. The amplitude of this disparity can be controlled by varying ϕ. However the angle between the two arms can not be too small. The distance between the optical centers (baseline) must be large enough to guarantee accurate 3D motion/trajectory recovery. Another important issue is that to each ϕ there are two different disparity values depending on the chosen α_p solution. The system must decide which is the most suitable choice for both ϕ and α_p in each circumstance. This requires a complex control strategy. The problem can be simplified by assuming $\phi = \pi$. By assuming this, the baseline is maximized (it becomes equal to $2B$), the α_p solutions are supplementary (we only need to compute one) and the corresponding disparity amplitudes are the same (see Fig. 1(4)). This is a trade-off solution. By fixating the angle between the arms, the disparity amplitude can no longer be controlled. But even when ϕ is varied, the amplitude is always limited to a certain range. And this range in only significant when the distance of the targets is in the same order of magnitude of the arm length B.

$$\tan(\alpha_p) = -\frac{(x_1^2 + z_1^2 - B^2)z_2 - (x_2^2 + z_2^2 - B^2)z_1}{(x_1^2 + z_1^2 - B^2)x_2 - (x_2^2 + z_2^2 - B^2)x_1} \quad (4)$$

Since ϕ is constant and equal to π the cameras can be placed at the extremities of a rotating arm with length $2B$. Assume that the left camera is tracking $target_1$ and the right camera tracking $target_2$. Equation 4 is derived from equation 3. It gives the platform rotation angle α_p that guarantees symmetric disparities in both retinas. We proved that occlusion only occurs when the targets are aligned with the rotation center of the arm ($\theta_1 = \theta_2 + k.\pi$ with $(\rho_1, \theta_1), (\rho_2, \theta_2)$ the targets spherical coordinates). In this situation the platform is also aligned with the targets ($\alpha_p = \theta_1 \pm \pi/2$ and $\theta_1 = \theta_2$)

3 Comparing the use of a rotational platform against fixed pan-and-tilt units

In this section we compare the configuration that is being considered with the setup made up of two fixed pan-tilt units when both moving targets are in the field of view of both cameras. The distance between the fixed units is $2B$ (equal to the length of the rotating arm). Fig.3(L) represents the experiment. $Target_1$ and $target_2$ move along a rectilinear path. The positions are given by two sinusoids with the same angular velocity ($w = 0.2\pi rad/s$) and with a phase difference of π.

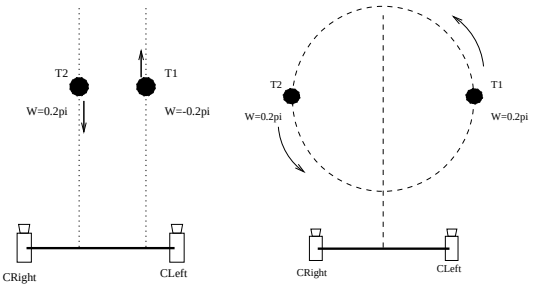

Figure 3: Left:Targets moving on linear paths. Right:Targets moving on a circular path

Fig.4(1) shows the evolution of the α_p angle for the dynamic configuration that guarantees symmetric disparity in the two images. Fig.4(2)(3) compares what happens in the left and right cameras in the case of a rotating arm and when the pan-tilt units do not move. Assume that the left camera is verged on $target_1$ and the right camera fixated on $target_2$. Notice that occlusion occurs at different time instants for both the left and right images in the case of the fixed pan-tilt units. The configuration with a rotating arm avoids this situation and 3D trajectory recovery is always possible. The fact that the two targets cross in the image also complicates the tracking process. In that case target position and velocity estimation is more difficult, and additional processing is required to identify which target to track after the crossing. The dynamic configuration (with the rotating arm) avoids those situations. The tracking algorithm can use the fact that the left camera tracks the target projected on the left side of the image and right camera tracks the target located on the right side of the image.

4 System control

In the previous sections the advantages of using a rotating platform and the best rotation angles were discussed. In this section we discuss how to control the system using visual information.

$$\Delta\alpha_p = \arctan(-2.\frac{\Gamma}{\Lambda}) \quad (5)$$

$$\Gamma = \tan(\theta_{L2} - \theta_{R2}) - \tan(\theta_{L1} - \theta_{R1})$$
$$\Lambda = (\tan(\theta_{L2}) + \tan(\theta_{R2})).\tan(\theta_{L2} - \theta_{R2}) -$$

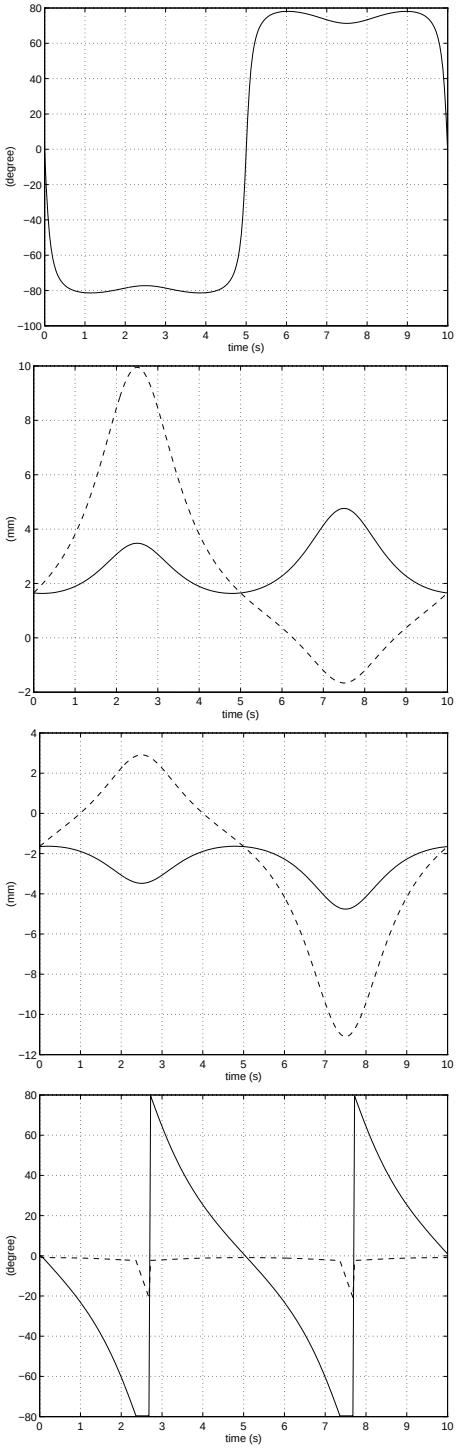

Figure 4: 1: Platform rotation angle. 2: Target position in the left image. Fixed (- -) and dynamic configuration(-). 3: Target position in the right image. Fixed (- -) and dynamic configuration(-). 4: Experiment 2. Platform rotation angle control. Evolution of angular position (-) and error (- -).

$$-(\tan(\theta_{L1}) + \tan(\theta_{R1})).\tan(\theta_{L1} - \theta_{R1})$$

$$\frac{d\Delta\alpha_p}{dt} = \frac{4.\Upsilon}{\sin(2(\theta_{L2} - \theta_{R2})).\Psi} \tag{6}$$

$$\begin{aligned}\Upsilon = &\quad \frac{d\theta_{L2}}{dt} - \frac{d\theta_{L1}}{dt} + \frac{d\theta_{R1}}{dt} - \frac{d\theta_{R2}}{dt} \\ \Psi = &\quad \tan(\theta_{R1}) - \tan(\theta_{R2}) - \tan(\theta_{L2}) + \tan(\theta_{L1})\end{aligned}$$

From equations 4 we derived equation 5. This expression gives the angular position error of the platform as a function of the targets angular positions in the camera coordinate system (see Fig 2). Equation 6 specifies the angular velocity variation of the platform as a function of the targets angular velocities. It is derived by differentiating equation 5 assuming $\Delta\alpha_p = 0$ and equal target disparity in both retinas.

$$\theta_{Ci} = \arctan(\frac{x_{img}^{Ci}}{f}) + \beta_C \tag{7}$$

$$\frac{d\theta_{Ci}}{dt} = \frac{dx_{img}^{Ci}}{dt}.\frac{\cos^2(\theta_{Ci} - \beta_C)}{f} + \frac{d\beta_C}{dt} \tag{8}$$

Consider camera C ($C = L, R$) and the corresponding coordinate system (see Fig 2). Equations 7 and 8 give $target_i$ ($i = 1, 2$) angular position and velocity as a function of camera motion and target motion in image. Camera angular motion can be measured using motor encoders. Thus, combining equations 5 to 8 with visual information, we are able to control platform rotation both in position and velocity. The extraction of information from images is simplified by the symmetric disparity.

Each camera is mounted on a pan and tilt unit that aims at keeping the corresponding target projection in the center of the image. Consider that θ_{Ci} is the target angular position in the fixed camera coordinate system and β_C the pan rotation angle. To keep the target in the center of the image β_C must be equal to θ_{Ci}. The tracking position error can be measured using equation 7. To achieve high speed tracking it is important to use both position and velocity control. Equation 8 gives the velocity error by measuring the velocity of the target in the image. Notice that the velocity induced in the image by pan motion must compensate for both the velocity induced by target motion and platform rotation [7].

Occlusion only occurs when the targets are aligned with the rotation center of the arm. Consider the experiment depicted in Fig 3(L). The targets move along a circular path with the same angular velocity.

They start their movements at opposite positions on the path. Occlusion occurs whenever $\theta_1 = \theta_2 = \pi/2$.

Assume that α_p takes values between $-\pi/2$ and $\pi/2$ (the supplementary solutions are excluded). When the disparity (equal in both images) goes under a threshold the arm stops until the target images cross each other. It only restarts rotating when segmentation becomes possible again. Consider that the platform rotation is controlled in position using equation 5. Fig.4(4) shows the evolution of α_p and $\Delta\alpha_p$. Notice that when the platform reaches -80deg rotation stops and the error increases until $\alpha_p + \Delta\alpha_p \epsilon [-80, 80]$. Due to the fact that supplementary solutions were excluded a sudden rotation inversion can be observed . Velocity control can be used to smooth this transition. The drawback of this approach is that when the platform stops rotating disparity is not equal in the two images. The left camera is programmed to track the leftmost target in the image and correspondingly the right camera tracks the rightmost target. Thus when targets cross in the image the cameras switch the targets they are tracking.

5 Target segmentation in image

To control the system we need to estimate both targets positions and velocities in the two images.

$Target_1$ is nearly at the center of left image. The same happens with $target_2$ and the right image. If the average target distance is significantly higher than B (arm length), the velocity induced in the image by system motion (egomotion) is almost constant for a neighborhood of pixels around the image center(see equation 8). Consider the central sub-image of two frames grabbed sequentially . The first sub-image is compensated for the egomotion (estimated using the motor encoders). A difference mask is obtained by subtracting the current image from the previously compensated. The velocity induced in image by target motion is estimated applying differential flow. The difference mask selects the pixels that are used in the computation. By doing this the background motion is compensated and the speed and robustness of the process is increased. A constant velocity model was used in the experiments. However other models that can be complemented by flow segmentation can lead to better results. The egomotion component due to arm rotation is added to obtain the correct velocity value to be used in equation 8. Target position is estimated as the average location of the set of difference points with non-zero brightness partial derivatives with respect to X and Y in the current frame.

$Target_1$ motion in right image and $target_2$ motion in left image are measured using correlation. If the arm length B is considerable smaller than the target distance, the correlation mask can be updated by segmenting the target projection at the center of the other image. This is particularly useful when the system is tracking non-rigid bodies.

6 Conclusions

The proposed configuration aims at dealing with specific instances of tracking of two objects, namely when they are simultaneously visible by the cameras. The main criterion used is the requirement that target disparities in both images are equal. This simplifies motion segmentation and allows 3D trajectory recovery. As a result of the specific configuration it can also be proved that occlusions are minimized. The configuration consists on two active cameras that move along a circular trajectory. Motion segmentation uses the fact that, as a result of the geometry, image projections of the targets are symmetrical. By using a third wide-angle camera, target selection becomes possible and the system can operate in environments where several moving targets are present.

References

[1] J. L. Crowley. Coordination of action and perception in a surveillance robot. *IEEE Expert*, 2, November 1987.

[2] T. Kanade, R.T. Collins, A.J. Lipton, P. Burt, and L. Wixon. Advances in cooperative multi-sensor video surveillance. In *DARPA98*, pages 3–24, 1998.

[3] Y. Guo, G. Xu, and S. Tsuji. Understanding human motion patterns. In *ICPR94*, pages B:325–329, 1994.

[4] N. Oliver, B. Rosario, and A. Pentland. A bayesian computer vision system for modeling human interactions. In *ICVS99–First Int. Conf. on Computer Vision Systems*, pages 255–272, 1999.

[5] J. Ferryman, S. Maybank, and A. Worrall. Visual surveillance for moving vehicles. In *Proc. of the IEEE Workshop on Visual Surveillance*, pages 73–80, 1998.

[6] J. Batista, P. Peixoto, and H. Araujo. Real-time active visual surveillance by integrating peripheral motion detection with foveated tracking. In *Proc. of the IEEE Workshop on Visual Surveillance*, pages 18–25, 1998.

[7] J. Barreto, P. Peixoto, J. Batista, and H. Araujo. Improving 3d active visual tracking. In *ICVS99–First Int. Conf. on Computer Vision Systems*, pages 412–431, 1999.